Mobile Health

Audio Signal and Health (2)

Cecilia Mascolo

UNIVERSITY OF CAMBRIDGE

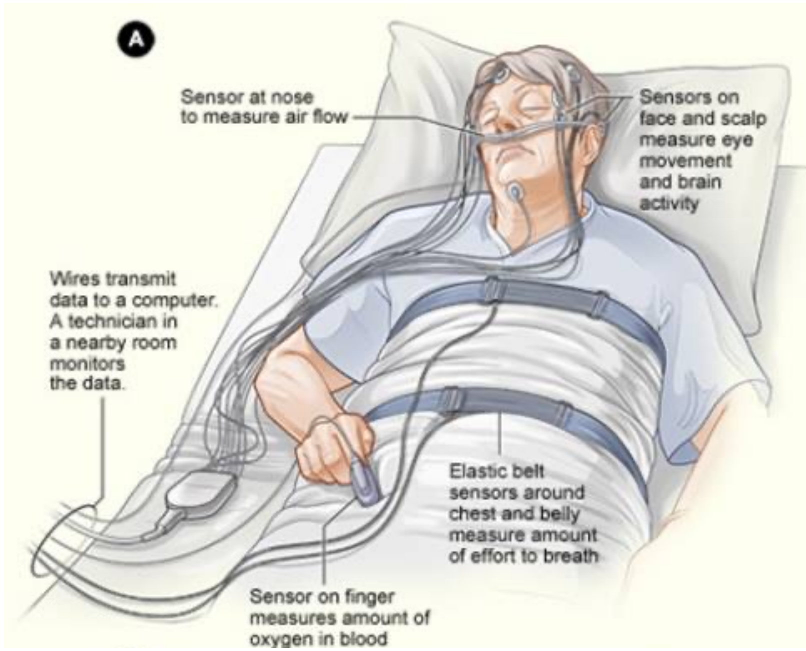
Sleep Stages Classification with Audio

- During sleep (in contrast to wakefulness) there is an increase of upper airway resistance due to decreased activity of the pharyngeal dilator muscles, which is reflected **by amplification of air-pressure oscillations during breathing**. These air-pressure oscillations are perceived as breathing sounds during sleep.
- REM (rapid-eye movement), N(on)REM, and wakefulness are associated with lack of, some, and considerable body movement.
- Breathing pattern is more periodic and consistent in deep NREM sleep compared to REM and wakefulness

Dafna, E., Tarasiuk, A. & Zigel, Y. Sleep staging using nocturnal sound analysis. *Sci Rep* **8**, 13474 (2018).

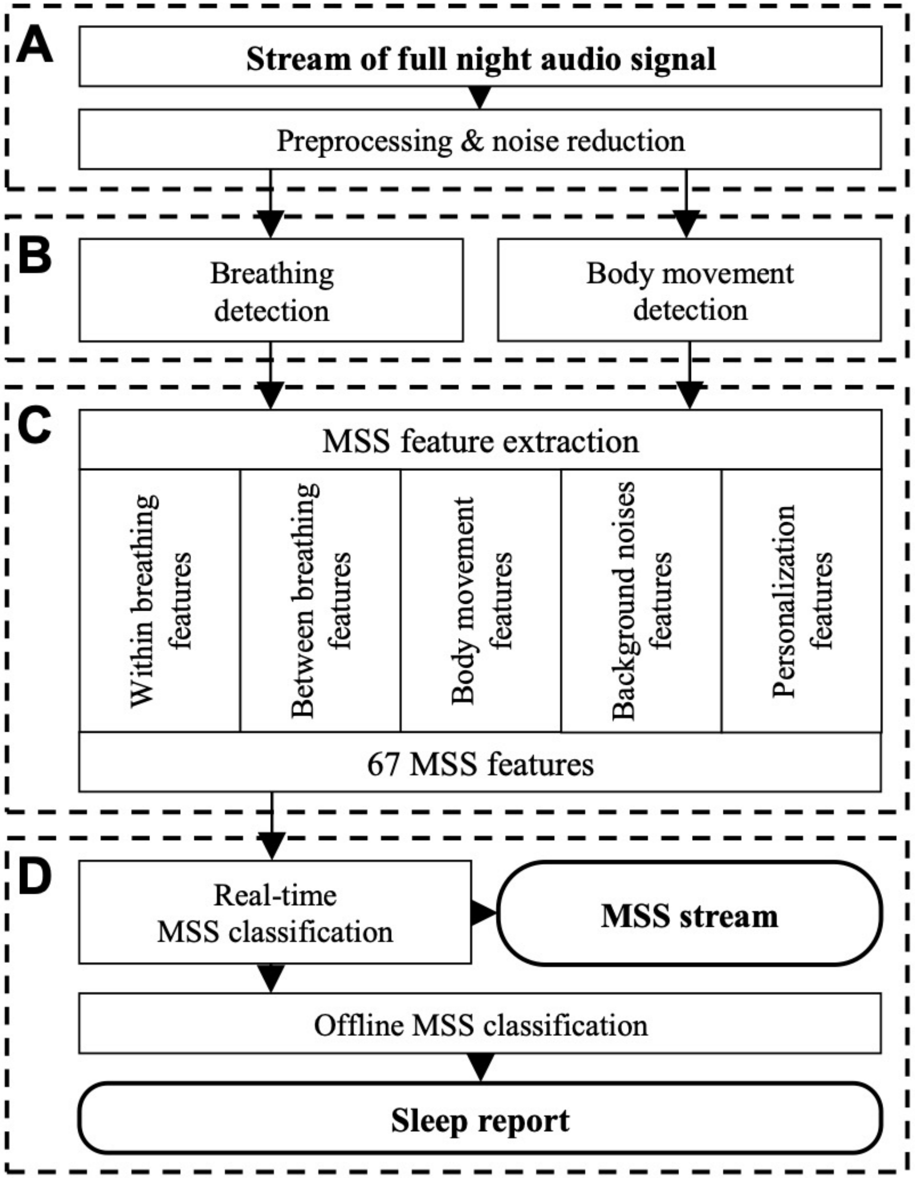
Audio

- Microphone on the bed: (Edirol R-4 pro, Bellingham, WA, USA) with a directional microphone (RØDE, NTG-1, Silverwater, NSW, Australia) was placed at a distance of one meter above the subject’s head and used for acquiring the audio signals.
- Polisomnography (PSG) for ground truth

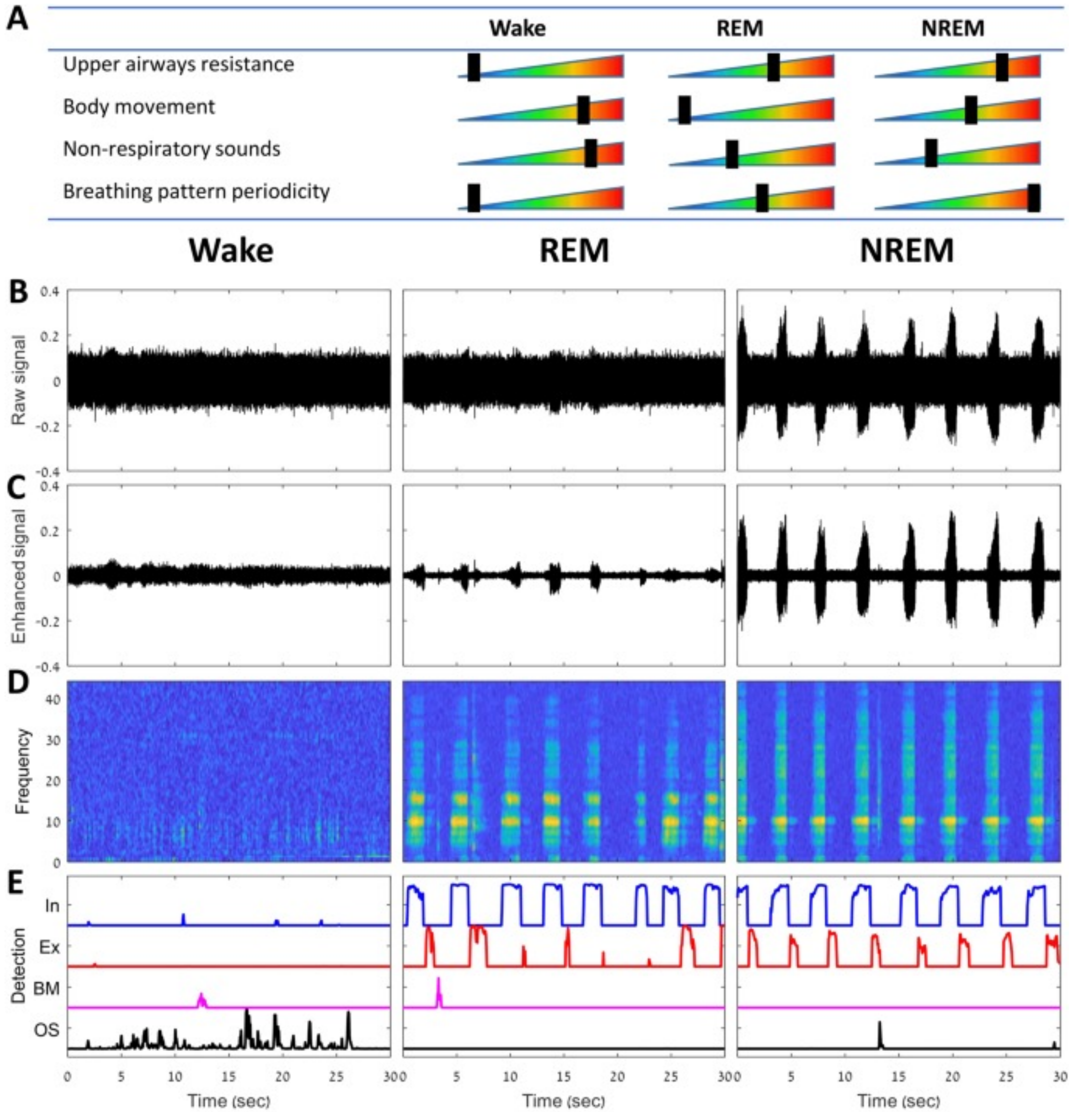


UNIVERSITY OF CAMBRIDGE

Detection of Macro Sleep Stages (MSS)



UNIVERSITY OF CAMBRIDGE



Raw sound

Preprocessed

Spectrogram

Inhalation (blue), Exhalation (red), body movement (pink) and other (black)

UNIVERSITY OF CAMBRIDGE

Within Breathing Features

- During sleep, airways resistance is higher than during wakefulness, hence breathing efforts become greater, which translates into several factors including l**ouder breathing** sounds, prolonged **breathing duration**, and **different vocal sounds** (snores).

		count	importance
A. Within breathing features (WB)	**Feature code**	**33**	**0.270**
Detection score of inspiration (μ,σ)	WB_DI	2	0.093
Detection score of expiration (μ,σ)	WB_DE	2	0.048
Detection score of respiration (μ,σ)	WB_DR	2	0.037
Duration inspiration (μ,σ)	WB_DurI	2	0.075
Duration expiration (μ,σ)	WB_DurE	2	0.024
Stationarity inspiration (μ,σ)	WB_SI	2	0.013
Stationarity expiration (μ,σ)	WB_SE	2	0.009
Sound intensity inspiration (μ,σ)	WB_SII	2	0.044
Sound intensity expiration (μ,σ)	WB_SIE	2	0.009
Sound intensity inspiration top 1% (μ,σ)	WB_SII01	2	0.027
Sound intensity expiration top 1% (μ,σ)	WB_SIE01	2	0.053
Entropy inspiration (μ,σ)	WB_EI	2	0.045
Entropy expiration (μ,σ)	WB_EE	2	0.008
Frequency centroid inspiration (μ,σ)	WB_FCI	2	0.031
Frequency centroid expiration (μ,σ)	WB_FCE	2	0.036
Frequency bandwidth (resp., insp., expi.)	WB_FB	3	0.009

UNIVERSITY OF CAMBRIDGE

Between Breathing Features

- Alternations in ventilation may affect fundamental respiration factors such as respiratory cycle period, respiratory duty cycle, and respiration consistency, and can be measured using sound analysis. These respiration factors are most likely to have more substantial variability during REM as opposed to NREM.

B. Between breathing features (BB)		**12**	**0.267**
Respiration duty cycle	BB_DCR	1	0.026
Inspiration duty cycle	BB_DCI	1	0.058
Expiration duty cycle	BB_DCE	1	0.020
Respiration cycle period (μ,σ)	BB_RCP	2	0.033
Respiration cycle period consistency	BB_RCPC	1	0.068
Respiration cycle periods fourth-order curve	BB_RCPfit	5	0.023
Breathing Count	BB_BC	1	0.006

UNIVERSITY OF CAMBRIDGE

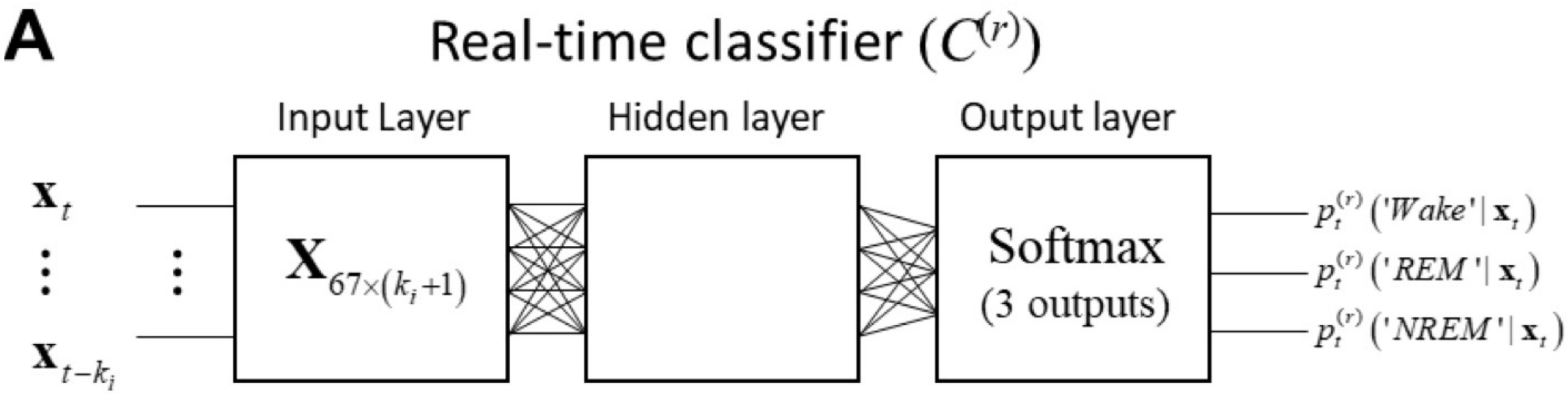
Body Movement Features

- Wakefulness is accompanied by relatively greater body movement, compared to NREM, while during REM sleep body movement should be absent by definition.

C. Body movement features (BM)		**10**	**0.054**
Body movement average score	BM_AS	1	0.002
Body movement overall score percentiles	BM_OS	7	0.017
Sound intensity body movement (all curve)	BM_SI	1	0.007
Sound intensity body movement 10% (all curve)	BM_SI01	1	0.038

UNIVERSITY OF CAMBRIDGE

Real time Classification

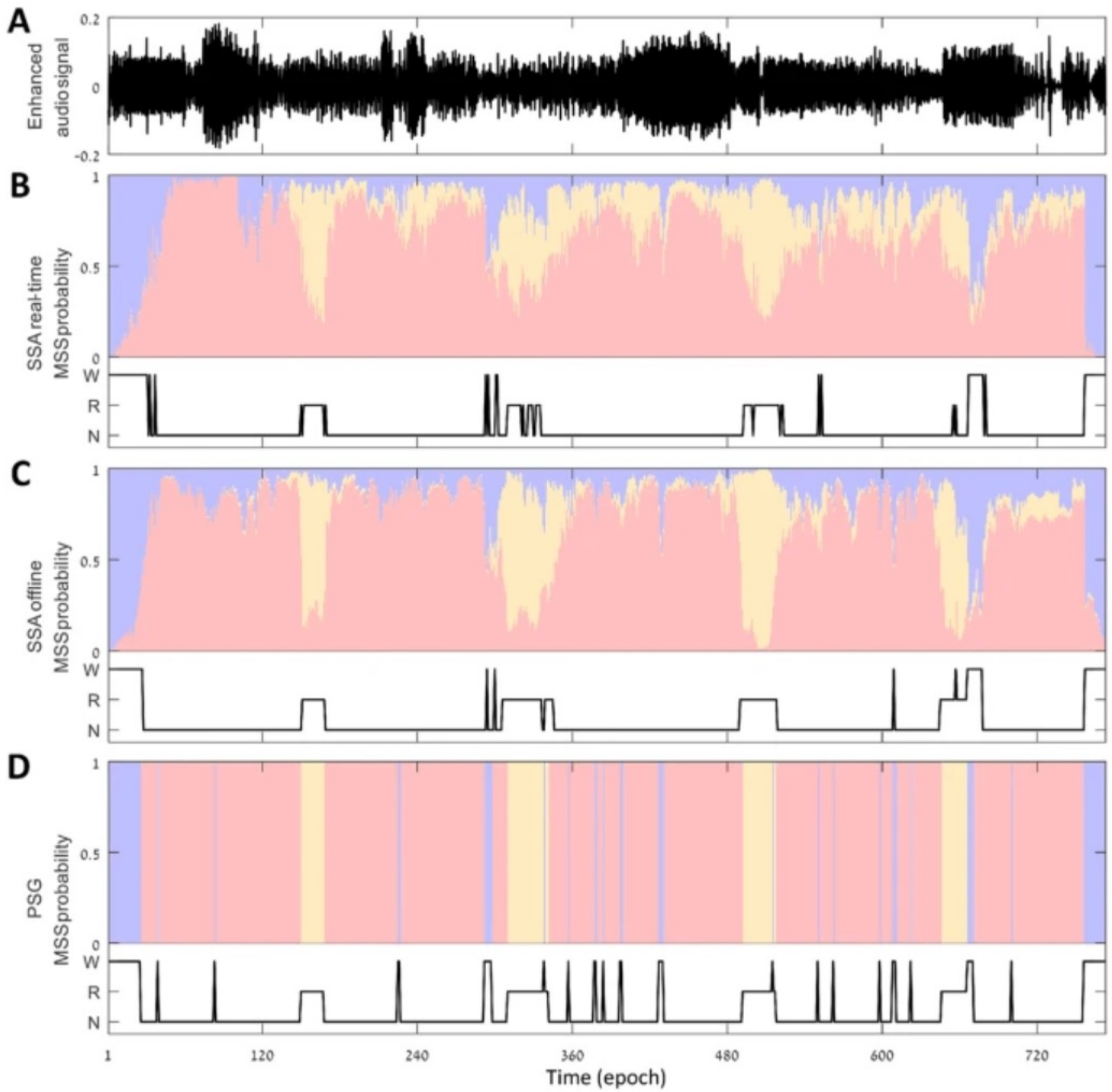


UNIVERSITY OF CAMBRIDGE

Results

One Subject

Blue= wake
Orange= REM sleep
Red= Non-REM sleep



UNIVERSITY OF CAMBRIDGE

Coronary Heart Disease and Voice

- In Coronary Heart Disease, plaque builds in arteries (which carry oxygen to the heart) and restricts flow.

- These changes can induce respiration changes, irregular breathing and increased muscle tension in the vocal tract.

- Participant’s voice while sustaining vowels was analyzed.

UNIVERSITY OF CAMBRIDGE

Feature: Average Fundamental Frequency

- Fundamental frequency (FF) is the rate of vocal fold vibration
  - FF: lowest frequency of a periodic waveform.
- Average all the extracted fundamental frequencies period by period.

Segment of a speech signal, with the period length *L*, and fundamental frequency *F0=1/L*.

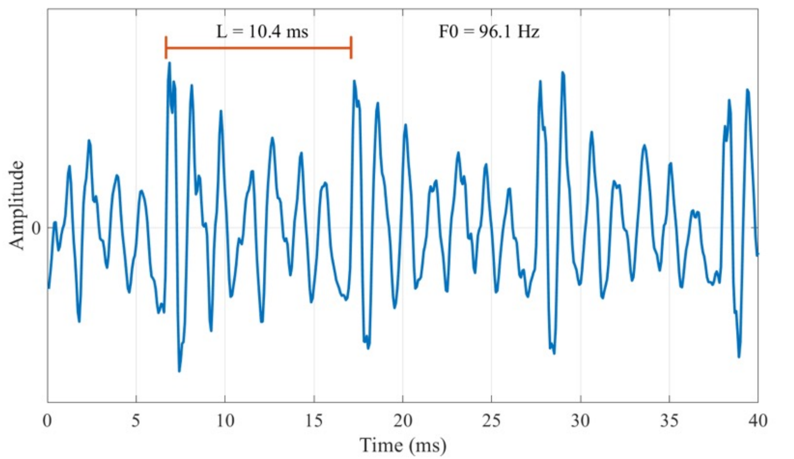
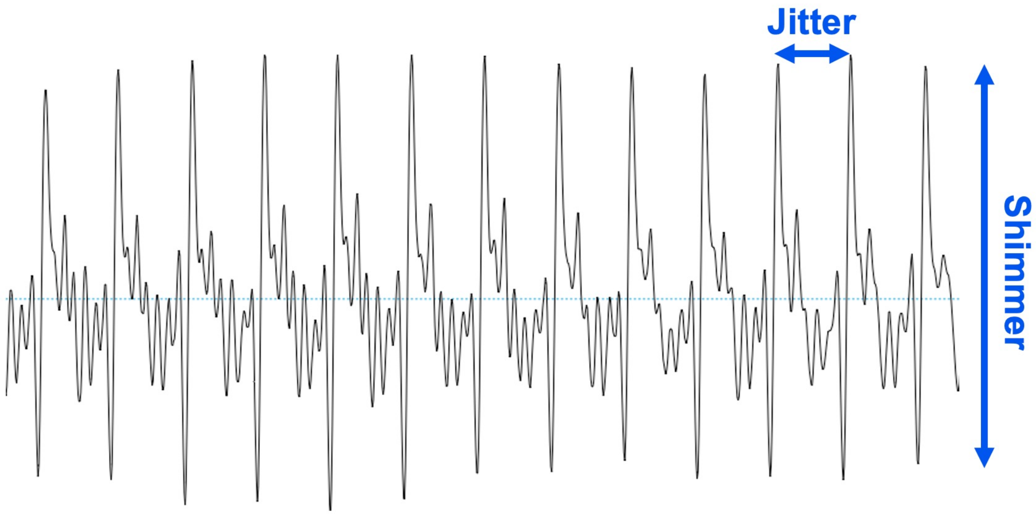


Figure from https://wiki.aalto.fi/pages/viewpage.action?pageId=149890776

UNIVERSITY OF CAMBRIDGE

Jitter and Shimmer

- Amount of variation in period length and amplitude are known respectively as *jitter* and *shimmer*.
- They are perceived as roughness, breathiness, or hoarseness in a speaker's voice.



UNIVERSITY OF CAMBRIDGE

Figure from https://wiki.aalto.fi/display/ITSP/Jitter+and+shimmer

Features: Absolute Jitter

- Absolute Jitter is the period to period variability of the pitch period
- Jitter in essence measures the changes in distance between peaks

$$\text{Jita} = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| T^{(i)} - T^{(i+1)} \right|$$

Feature: Shimmer

- Measures the differences between amplitudes of the max peaks in periods

UNIVERSITY OF CAMBRIDGE

Results (male group)

Parameters	Control Group (mean ± SD)	CHD Group (mean ± SD)
Jita (µsec)	116.56±41.09	68.45±27.88
Jitt (%)	1.35±0.509	0.86±0.41
RAP (%)	0.80±0.30	0.54±0.30
PPQ (%)	0.81±0.30	0.53±0.31
sPPQ (%)	1.19±0.52	0.76±0.28
ShdB (dB)	0.73±0.25	0.45±0.16
Shim (%)	8.056±2.59	4.98±1.59
APQ (%)	5.87±1.76	3.70±1.14
sAPQ (%)	8.69±3.23	6.32±2.30

UNIVERSITY OF CAMBRIDGE

Parkinson’s

- **Parkinson's disease** is a brain disorder that leads to shaking, stiffness, and difficulty with walking, balance, and coordination.
- Hypokinetic dysarthria (HD) occurs in 90% of Parkinson’s disease (PD) patients.
- HD is characterized by rigidity, bradykinesia, and **reduced muscular control of the larynx**, articulatory organs, and **other physiological support mechanisms of human speech production**. The following speech flaws have been observed**: increased acoustic noise, reduced intensity of voice, harsh and breathy voice quality, increased voice nasality, monopitch, monoloudness, and speech rate disturbances**.

UNIVERSITY OF CAMBRIDGE

Parkinson’s diagnosis via voice: Shimmer works

Vowel	Feature	ρ	MI	p	ACC [%]	SEN [%]	SPE [%]	TSS
a (s)	F_2 (99p)	–0.0219	0.7540	0.8029	65.41	66.67	63.27	1.65
e (s)	BW_2 (1p)	–0.0045	0.5826	0.9609	68.42	69.05	67.35	1.71
i (s)	IMF-SNR$_{\mathrm{TKEO}}$ (ir)	–0.0865	0.3564	0.3216	68.42	72.62	61.22	1.68
o (s)	IMF-SNR$_{\mathrm{SE}}$ (1p)	0.0946	0.5631	0.2781	68.42	72.62	61.22	1.68
u (s)	IMF-SNR$_{\mathrm{SEO}}$ (std)	–0.0568	0.6674	0.5152	67.67	67.86	67.35	1.70
a (l)	IMF-SNR$_{\mathrm{SEO}}$ (1p)	0.0897	0.3127	0.3037	63.16	64.29	61.22	1.62
e (l)	IMF-GNE (median)	–0.0747	0.4386	0.3920	63.91	63.10	65.31	1.64
i (l)	IMF-NSR$_{\mathrm{SE}}$ (1p)	0.0438	**0.7679**	0.6161	62.41	60.71	65.31	1.62
o (l)	F_0 (ir)	–0.0292	0.6948	0.7388	66.92	71.43	59.18	1.65
u (l)	IMF-GNE (99p)	–0.0309	0.2310	0.7247	68.42	71.43	63.27	1.69
a (ll)	jitter (RAP)	–0.0568	0.4549	0.5152	69.92	73.81	63.27	1.70
e (ll)	IMF-NSR$_{\mathrm{RE}}$ (std)	–0.2911	0.6768	0.0008	66.92	66.67	67.35	1.69
i (ll)	IMF-CPP (median)	–0.1790	0.7071	0.0399	67.67	70.24	63.27	1.68
o (ll)	IMF-SNR$_{\mathrm{SE}}$ (1p)	–0.0345	0.6136	0.6935	62.41	69.05	51.02	1.55
u (ll)	IMF-NSR$_{\mathrm{SE}}$ (ir)	–0.2010	0.6654	0.0211	69.17	71.43	65.31	1.71
a (ls)	IMF-NSR$_{\mathrm{SE}}$ (median)	0.0930	0.7455	0.2865	64.66	67.86	59.18	1.63
e (ls)	IMF-NSR$_{\mathrm{TKEO}}$ (std)	–0.1636	0.6317	0.0605	66.17	63.10	71.43	1.69
i (ls)	shimmer (local, dB)	**–0.4064**	0.7633	**0.0000**	**72.18**	**75.00**	**67.35**	**1.75**
o (ls)	IMF-FD (median)	–0.2119	0.7276	0.0150	66.17	70.24	59.18	1.64
u (ls)	HNR (median)	0.2976	0.6768	0.0006	65.41	70.24	57.14	1.62

Z. Smekal, J. Mekyska, Z. Galaz, Z. Mzourek, I. Rektorova and M. Faundez-Zanuy, "Analysis of phonation in patients with Parkinson's disease using empirical mode decomposition," 2015 International Symposium on Signals, Circuits and Systems (ISSCS), 2015

UNIVERSITY OF CAMBRIDGE

OpenSmile Toolkit and Features

Audio features (low-level)

The following (audio-specific) low-level descriptors can be computed by openSMILE:

- Frame Energy
- Frame Intensity / Loudness (approximation)
- Critical Band spectra (Mel/Bark/Octave, triangular masking filters)
- Mel-/Bark-Frequency-Cepstral Coefficients (MFCC)
- Auditory Spectra
- Loudness approximated from auditory spectra
- Perceptual Linear Predictive (PLP) Coefficients
- Perceptual Linear Predictive Cepstral Coefficients (PLP-CC)
- Linear Predictive Coefficients (LPC)
- Line Spectral Pairs (LSP, aka. LSF)
- Fundamental Frequency (via ACF/Cepstrum method and via Subharmonic-Summation (SHS))
- Probability of Voicing from ACF and SHS spectrum peak
- Voice-Quality: Jitter and Shimmer
- Formant frequencies and bandwidths
- Zero and Mean Crossing rate
- Spectral features (arbitrary band energies, roll-off points, centroid, entropy, maxpos, minpos, variance (= spread), skewness, kurtosis, slope)
- Psychoacoustic sharpness, spectral harmonicity
- CHROMA (octave-warped semitone spectra) and CENS features (energy-normalised and smoothed CHROMA)
- CHROMA-derived features for Chord and Key recognition
- F0 Harmonics ratios

Heart Auscultation

One heartbeat consists of two sounds, commonly known as: “Lub” and “Dub”.

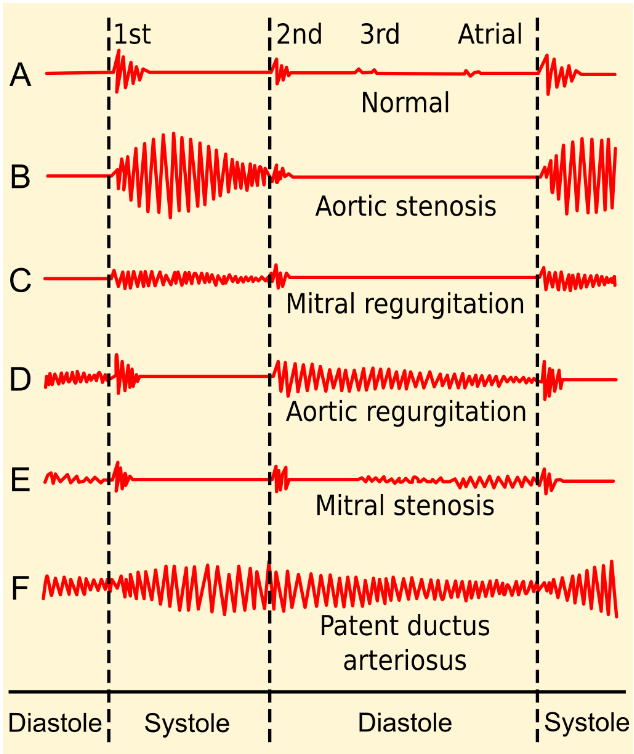
“Lub” = turbulence from closure of **mitral** and **tricuspid** valves
“Dub” = turbulence from closure of **aortic** and **pulmonic** valves

Trainee doctors from USA, UK, and Canada could only diagnose the heart pathology **correctly in 23% of cases** [1]

[1] S. Mangione, “Cardiac auscultatory skills of physicians-in-training: a comparison of three English- speaking countries,” Am. J. Med., vol. 110, no. 3, pp. 210–216, Feb. 2001.

UNIVERSITY OF CAMBRIDGE

Hear Pathology Diagnosis through Audio Data

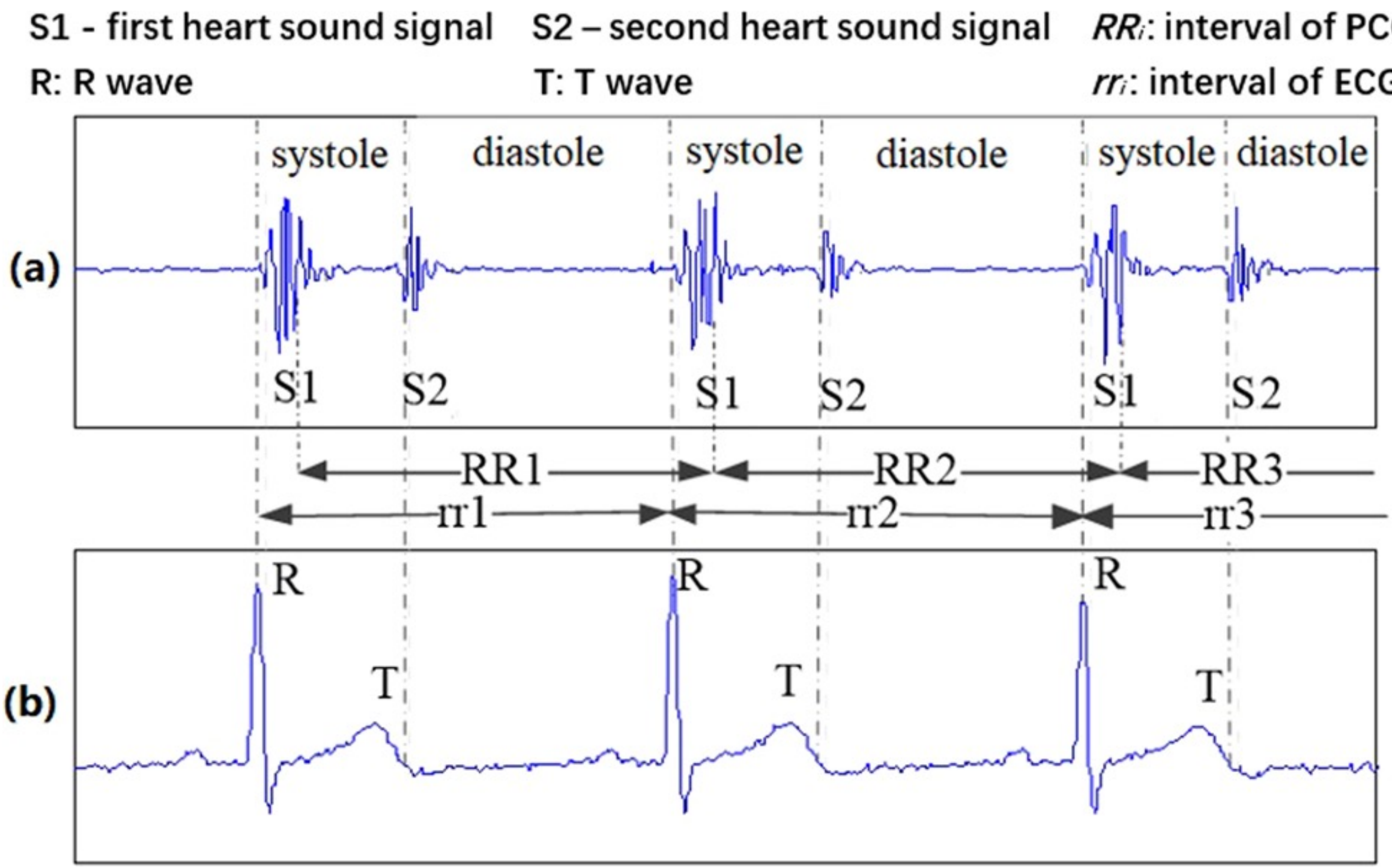


UNIVERSITY OF CAMBRIDGE

Alignment of ECG and Audio

PCG

ECG



FFT of S1 and S2 components

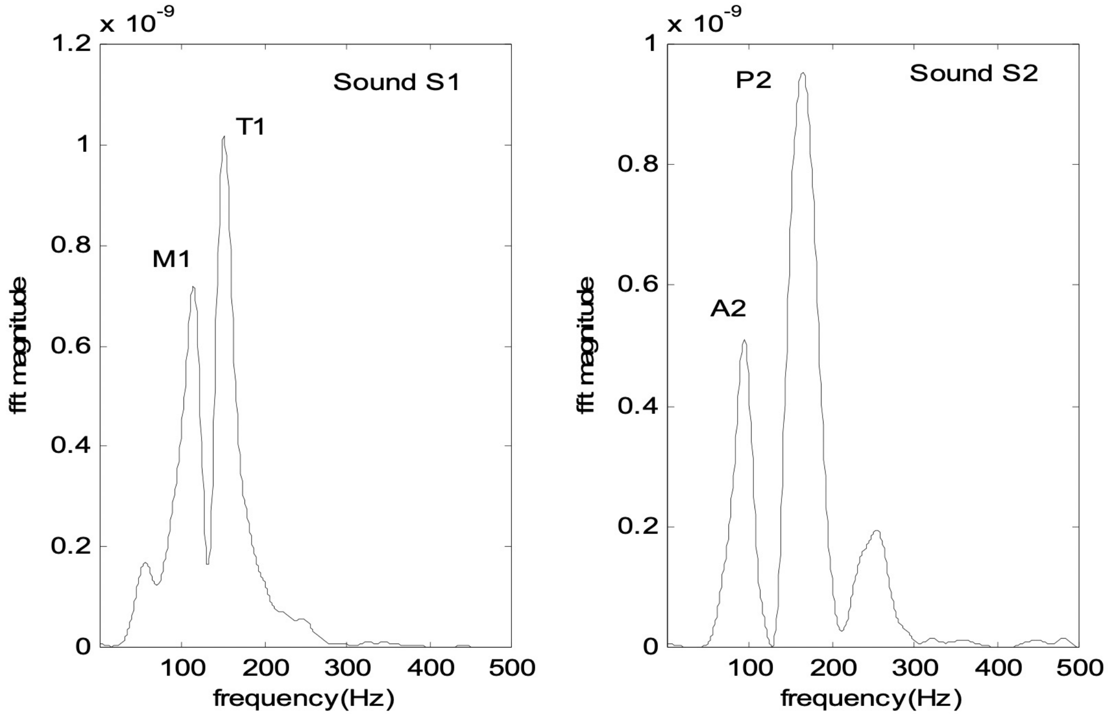


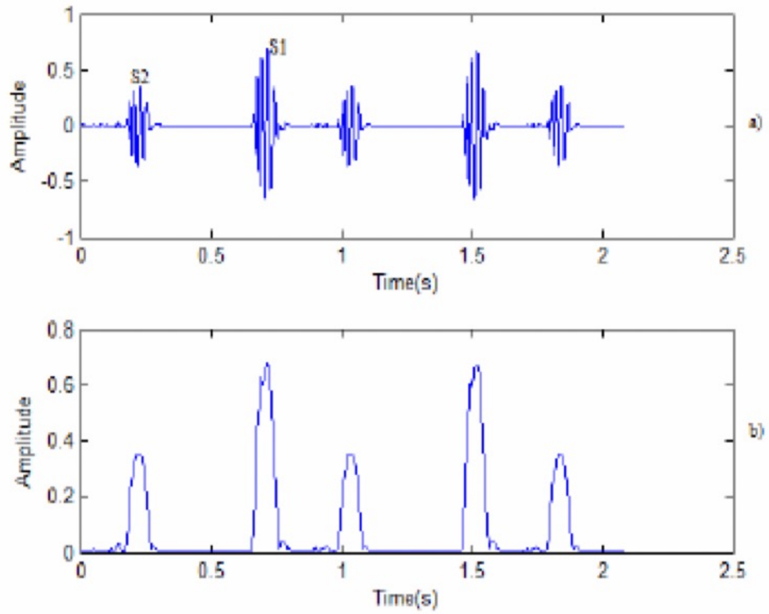
Figure from Debbal, S & bereksi reguig, Fethi. (2008). Frequency analysis of the heartbeat sounds. IJBSCHS. 13. 85-90.

UNIVERSITY OF CAMBRIDGE

Shannon Energy based Envelope Calculation

$$\text{Shannon Energy (t)} = -signal^2(\text{t}) * \log(signal^2(t))$$

$$\text{E}_{\text{avg}} = -\frac{1}{\text{N}}\sum_{\text{t}=1}^{\text{N}} \text{Shannon Energy(t)}$$

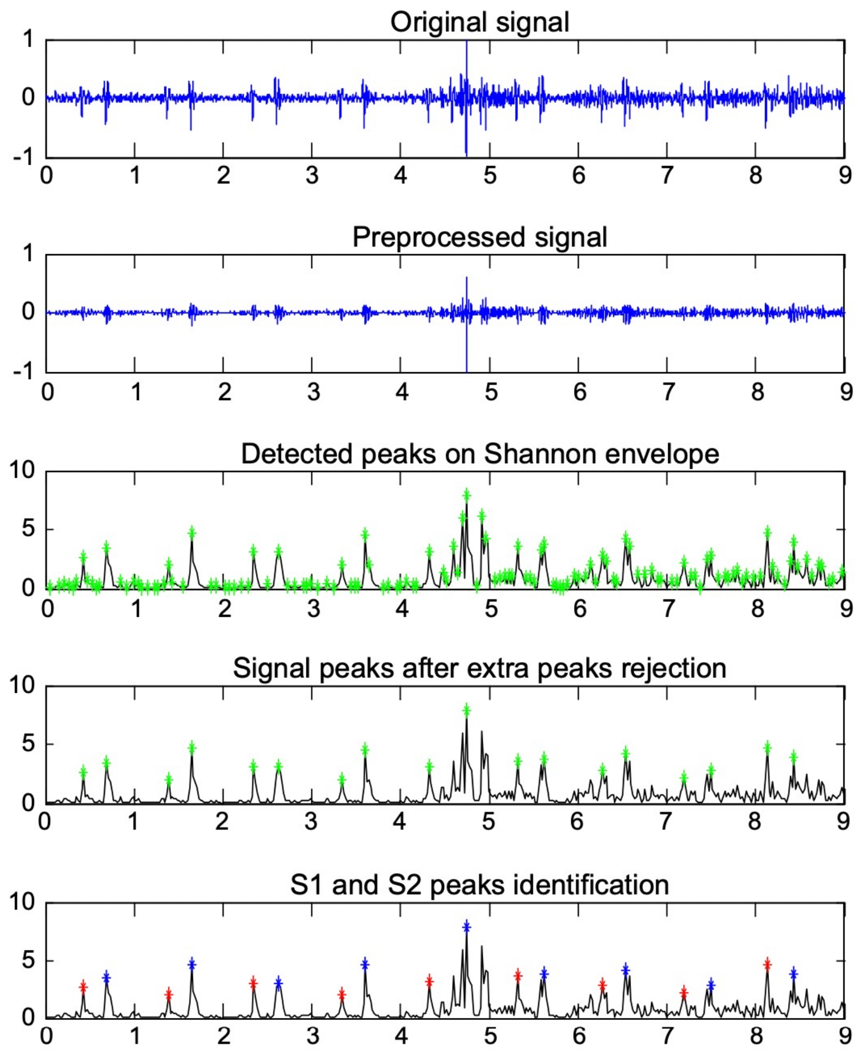


UNIVERSITY OF CAMBRIDGE

P. Sharma, S. Saha, S. Kumari (2018); Study and Design of a Shannon-Energy-Envelope based Phonocardiogram Peak Spacing Analysis for Estimating Arrhythmic Heart-Beat; Int J Sci Res Publ 4(9)

Shannon Energy based Peak Detection

- Rejection of extra peak is dataset dependent and based on peaks per time interval and their distance.



Chakir, Fatima et al. “Phonocardiogram signals processing approach for PASCAL Classifying Heart Sounds Challenge.” Signal, Image and Video Processing 12 (2018): 1149-1155.

Features…

[and KNN classifier]

Descriptor	Significance
T1	The interval between S1 and S2 peaks
T2	The interval between S2 and S1 peaks
F1	The sum of the amplitude variations between two successive samples of the signal during the period between S1 and S2 peaks divided by the length T1
F2	The sum of the amplitude variations between two successive samples of the signal during the period between S2 and S1 peaks divided by the length T2
Pw	The total original signal power
Es1	The standard deviation between S1 and S2 peaks
Es2	The standard deviation between S2 and S1 peaks
R	Takes the value 1 if there is an additional peak S1 or S2 out of rhythm; otherwise, it is equal to 0
L	Length of the signal
Zp	The zero crossing rate
Mn	The minimum amplitude of the signal
Mx	The maximum amplitude of the signal

UNIVERSITY OF CAMBRIDGE

Confusion Matrix of Classification

Table 3 Confusion matrix for Dataset A

	Normal	Murmur	Extra HS	Artifact	Total
Normal	10	1	1	2	14
Murmur	4	9	0	1	14
Extra HS	1	0	5	2	8
Artifact	2	1	0	13	16
Total	17	11	6	18	52

Table 2 Total error of the first PASCAL classifying heart sounds challenge found by our methodology and by other approaches

	Dataset A (s)	Dataset B (s)
ISEP/IPP Portugal	95.68	18.06
CS UCL	76.97	18.89
SLAC Stanford	28.2	19.11
UPD DCS Philippines	68.32	16.93
Our methodology	19.44	7.32

More general audio features

Feature Group	Description
Waveform	Zero-Crossings, Extremes, DC
Signal energy	Root Mean-Square & logarithmic
Loudness	Intensity & approx. loudness
FFT spectrum	Phase, magnitude (lin, dB, dBA)
ACF, Cepstrum	Autocorrelation and Cepstrum
Mel/Bark spectr.	Bands 0-N_{mel}
Semitone spectr.	FFT based and filter based
Cepstral	Cepstral features, e.g. MFCC, PLP-CC
Pitch	F_0 via ACF and SHS methods Probability of Voicing
Voice Quality	HNR, Jitter, Shimmer
LPC	LPC coeff., reflect. coeff., residual Line spectral pairs (LSP)
Auditory	Auditory spectra and PLP coeff.
Formants	Centre frequencies and bandwidths
Spectral	Energy in N user-defined bands, multiple roll-off points, centroid, entropy, flux, and rel. pos. of max./min.
Tonal	CHROMA, CENS, CHROMA-based features

E. Bondareva, J. Han, W. Bradlow, C. Mascolo. Segmentation-free Heart Pathology Detection Using Deep Learning. In Procs of Int. Conf. of the IEEE Engineering in Medicine and Biology Society. 2021.

Deep Learning Pipeline

- Of the 6K features:
  - Principal Component Analysis used to reduce features to ~500 feature vector
- A deep learning fully connected network is used (6 layers)

	Previous works					Our method	
	[3]	**[9]**	**[4]**	**[5]**	**[13]**	**SVM**	**DNN**
PN	0.70	0.77	0.71	**0.82**	0.77	**0.82**	0.81
PM	0.30	0.37	0.33	0.59	0.76	0.70	**0.96**
PE	0.67	0.17	**1.00**	0.18	0.50	0.20	0.50
Sens	0.19	0.51	0.14	0.49	0.34	**0.54**	0.47
Spec	0.84	0.59	0.90	0.66	0.95	0.77	**0.99**

E. Bondareva, J. Han, W. Bradlow, C. Mascolo. Segmentation-free Heart Pathology Detection Using Deep Learning. In Procs of Int. Conf. of the IEEE Engineering in Medicine and Biology Society. 2021.

Deep Learning

- Often generate vectors/matrices of features as input
- Construct a DNN architecture able to solve the task

Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning

Nicholas Cummins[a,*], Alice Baird[a], Björn W. Schuller[a,b]

[a] ZD.B Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany
[b] GLAM – Group on Language, Audio & Music, Imperial College London, UK

A R T I C L E I N F O

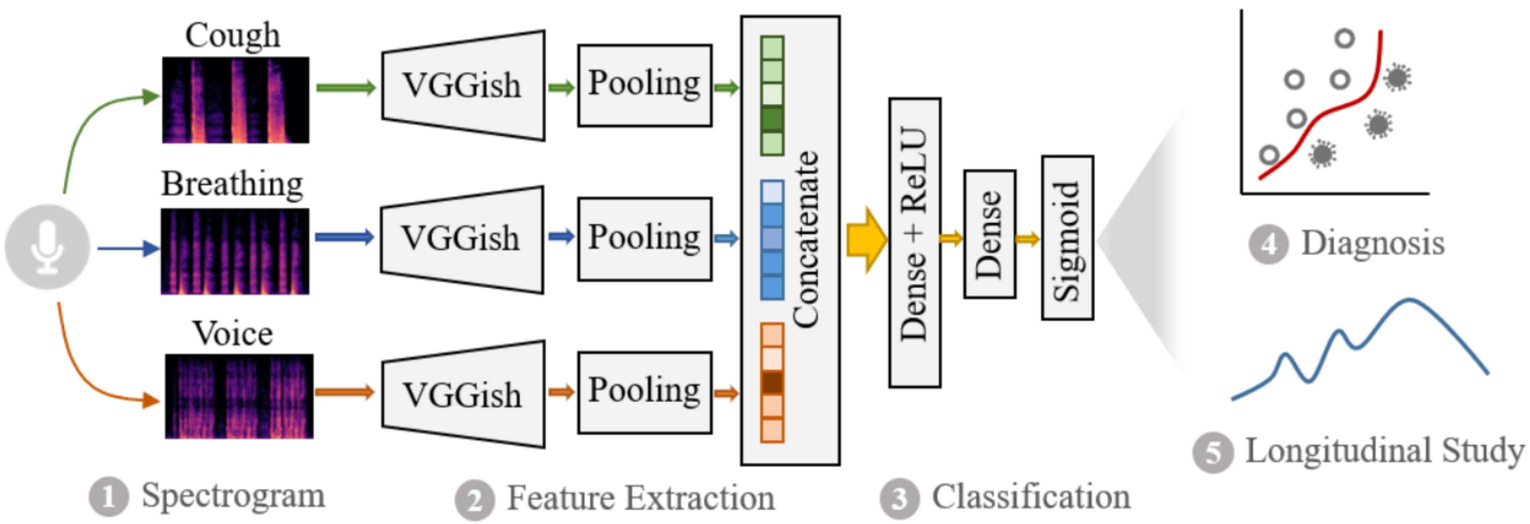
Keywords:
Speech

A B S T R A C T

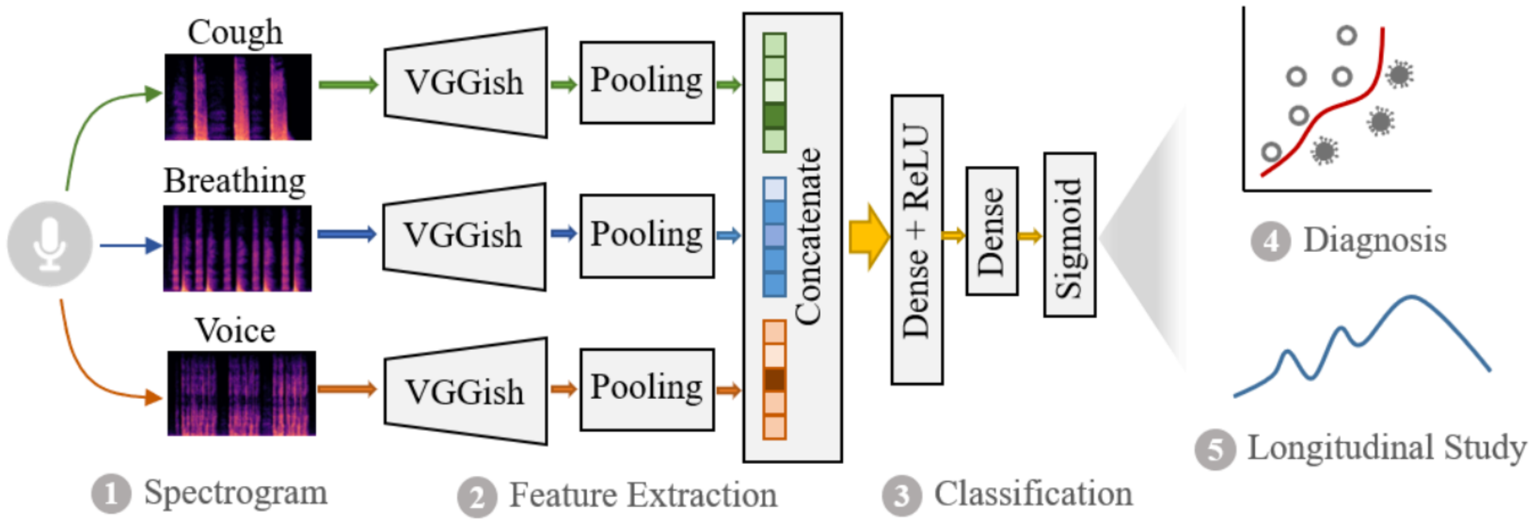
Due to the complex and intricate nature associated with their production, the acoustic-prosodic properties of a speech signal are modulated with a range of health related effects. There is an active and growing area of

Self Supervised and Transfer Learning

- Like for HAR, pretraining, self supervised and transfer learning are useful in audio analysis. Example of application of pretrained model:



COVID-19 Detection: pretrained audio model example



UNIVERSITY OF CAMBRIDGE

Questions

UNIVERSITY OF
CAMBRIDGE